



NTP
National Toxicology Program

The NIEHS/NTP-Perlegen Resequencing Project

F. M. Johnson, Ph.D.
Research Geneticist
National Institute of Environmental Health Sciences

NTP Board of Scientific Counselors
December 9-10, 2009





Project Significance and Relevance to NTP

- A. Provides a community resource for basic research.
- B. Sets the stage for investigating the genetic basis for differences in the toxic response.

Most toxicology testing utilizes very few genotypes, which tacitly assumes genotype makes little or no difference to test outcome.



NTP
National Toxicology Program

Key Project Personnel

Kelly Frazer, David Cox, Erica Beilharz, Perlegen Science



Molly Bogue, Jackson Laboratory



Mark Daly, Broad Institute



Eleazar Eskin, UCLA



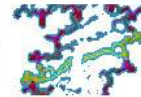


Phase 1:

- A. Resequence the genomes of 15 inbred mouse strains using the C57BL/6J sequence as a template.
- B. Organize the sequence by chromosome and chromosome location.
- C. Identify the variations in sequence that distinguish the strains.
- D. Analyze the haplotype structure of the strains and identify shared segments.
- E. Develop a website to make the methods and data available to the public.
- F. Submit the data to the National nucleotide data repositories, NCBI's dbSNP and Trace File Archive.
- G. Develop a research publication describing the results.



Single Nucleotide Polymorphism





The Results of Phase 1 were Published in Nature in 2007.

Frazer KA, Eskin E, Kang HM, Bogue MA, Hinds DA, . Beilharz EJ, Gupta RV, Montgomery J, Morenzoni MM, Nilsen GB, Pethiyagoda CL, Stuve LL, Johnson FM, Daly MJ, Wade CM and Cox DR. 2007. A sequence-based variation map of 8.27 million SNPs in inbred mouse strains. Nature 448(7157):1050-1053.

**8,266,653 SNPs identified among 2.57 billions base pairs,
1 SNP per 311 base pairs.**

Data incorporated into all major genomic data bases.

Cited in 100 research publications





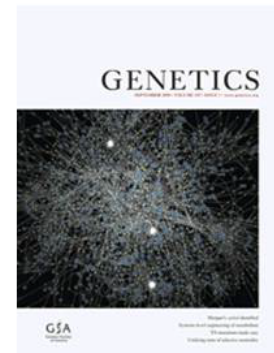
Phase 2:

- A. Impute the genotypes of 8.27 million SNPs in additional mouse strains.**
- B. Assist/facilitate integration of the SNP genotype data into the JAX phenome database and begin looking for genotype-phenotype associations.**
- C. Develop a research publication describing the results.**



Publication of the results of phase 2 is expected 2010.

Kirby, A, Kang HM, Wade CM, Cotsapas CJ, Kostem E, Han B, Rivas M, Bogue MA, Frazer KA, Johnson FM, Beliharz EJ, Cox DR, Eskin E and M Daly. A high density haplotype resource of 94 inbred mouse strains. Submitted to Genetics 2009.





15 Inbred Mouse Strains (*Mus musculus*), AKA The House Mouse

11 Classical Strains

129S1/SvImJ

A/J

AKR/J

BALB/cByJ

C3H/HeJ

DBA/2J

FVB/NJ

NOD/LtJ

BTBR T+tf/J

KK/HIJ

NZW/LacJ



BALB/c

4 Wild Derived Strains

CAST/EiJ (*M.m.castaneus*)

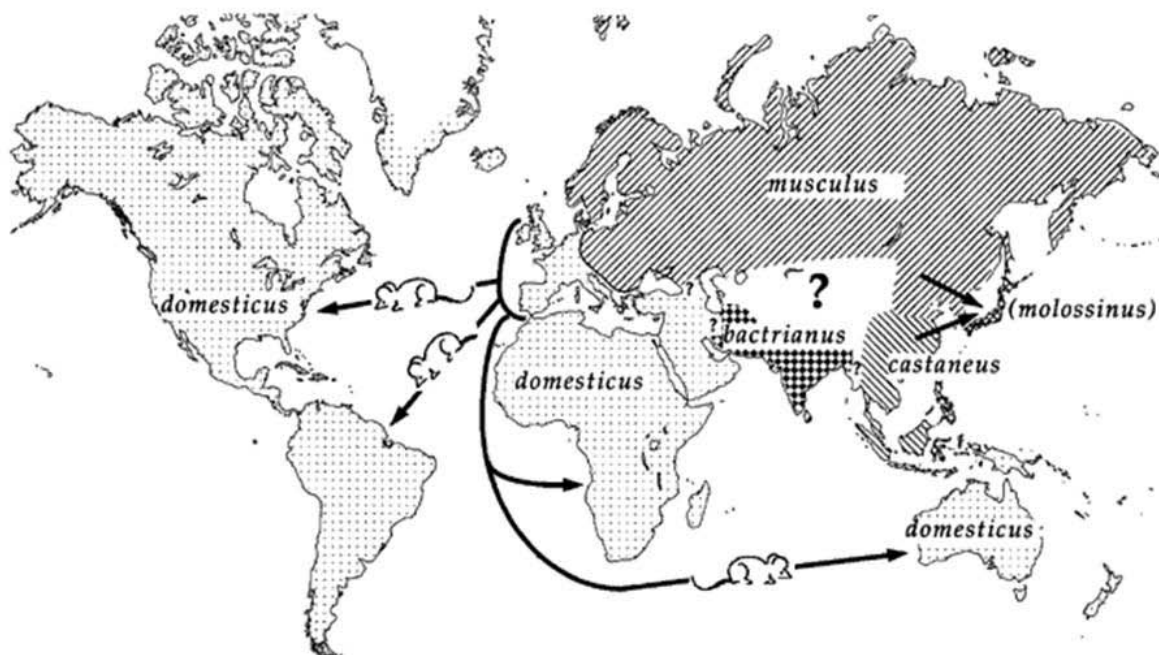
MOLF/EiJ (*M.m.molossinus*)

PWD/PhJ (*M.m.musculus*)

WSB/EiJ (*M.m.domesticus*)

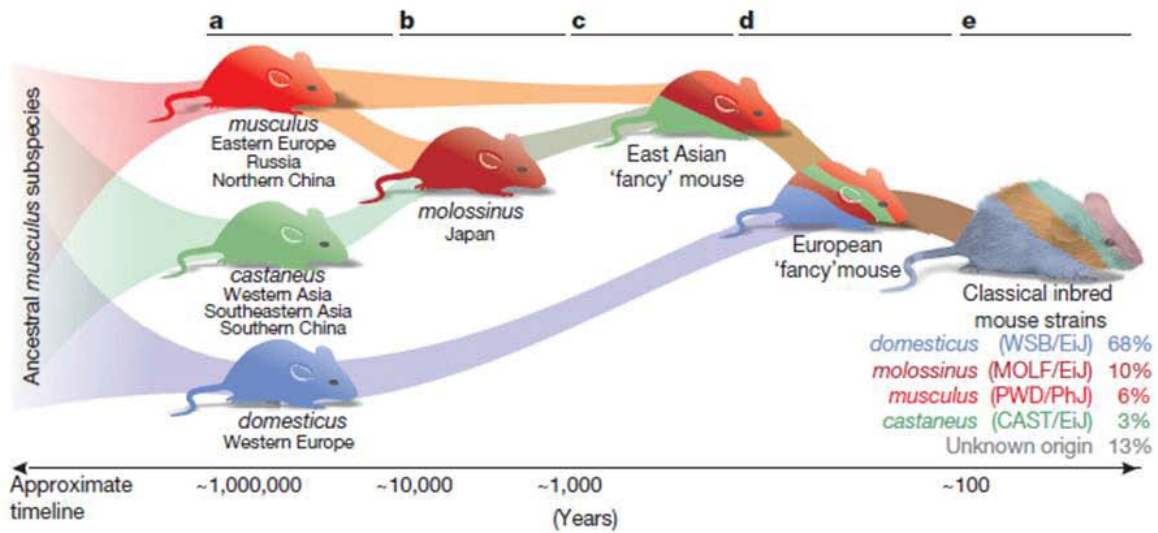




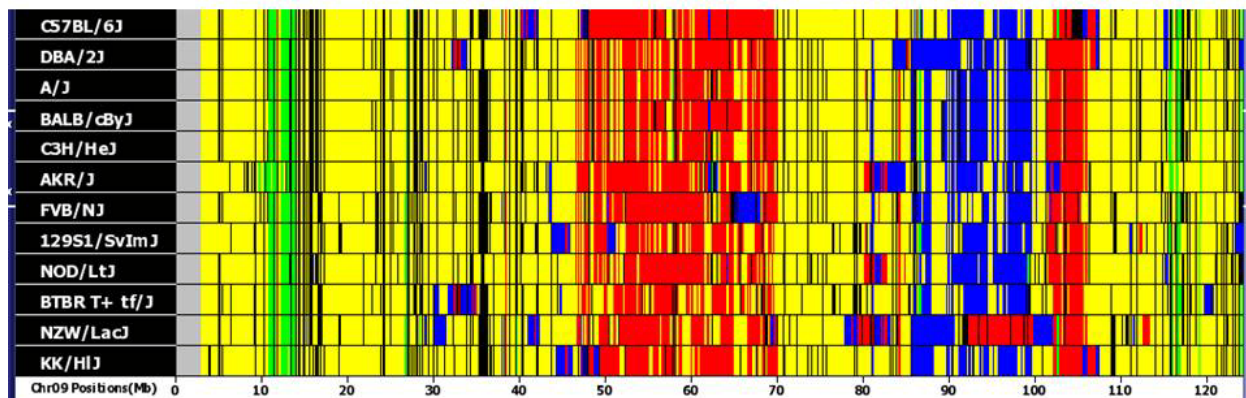


Geographical ranges of subspecies in the *M. musculus* group

In Mouse Genetics, Lee Siler, Oxford University Press 1995



Frazer K et al 2007. A sequence-based variation map of 8.27 million SNPs in inbred mouse strains. *Nature* 448(7157):1050-1053.



Haplotype Structure of Chromosome 9 in 12 Classical Inbred Mouse Strains



GeneChip Microarray Technology

25-mer oligonucleotide probes were generated from the known published genomic sequence of the C57BL/6J reference strain.

241,806 long-range PCR primer pairs were used to amplify the genomes of the 15 mouse strains. These “amplicons” ranged in length from 3 to 12 kb.

The probes were then used to interrogate the unknown sequence in the amplified regions of the 15 strains.

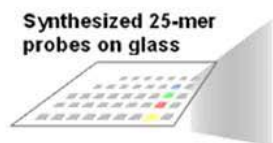
The process effectively examined the genomes of the 15 strains, one base at a time, in comparison with the B6 DNA sequence.



Reference DNA

```
...aaaatccatggttGcggttgtcacagg...
...tttaggtacaaCgcaacagtggtcc...
```

Synthesized 25-mer probes on glass



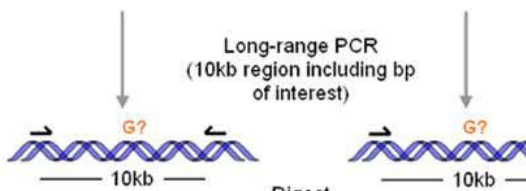
```
tttaggtacaaTgcaacagtggtcc
tttaggtacaaGgcaacagtggtcc
tttaggtacaaCgcaacagtggtcc
tttaggtacaaAgcaacagtggtcc
```

Samples

Mouse strain 1 genomic DNA

Mouse strain 2 genomic DNA

PCR products



Labeled DNA hybridized to high-density oligonucleotide array

T
G
C
A

T
G
C
A



NIEHS/Perlegen Project

2004	\$16,000,000	15 strains	\$1,060,000	1.5X Coverage per genome
------	--------------	------------	-------------	-----------------------------

In June, 2009, Illumina announced a personal genome sequencing service offering “complete” genome analysis of an individual person’s DNA for \$48,000 at 30X coverage.



In September, 2009, Complete Genomics’ Chairman, Clifford Reid, projected high-volume human genome sequencing at a cost as low as \$5,000 per 40x genome,



Complete Genomics

Powering large-scale
human genome studies



NTP
National Toxicology Program



Human Genome
Sequencing Center

Proposed sequencing 192 wild-derived inbred strains of the fruit fly, *D. melanogaster*.

A community research resource.

Human health related phenotypes like longevity, body size, and social behavior.

Inbred strains created from individuals captured in the wild.

Does not propose to examine genetic variation in the toxic response to different environmental chemicals, NTP's domain.

Mackay, T, Richards, S and Gibbs R. 2009. Proposal to sequence a *Drosophila* genetic reference panel: A community resource for the study of genotypic and phenotypic variation. http://flybase.org/static_pages/news/whitepapers/Drosophila_Genetic_Reference_Panel_Whitepaper.pdf.



Suggested Future Projects

Create a resource of 50 -100 newly derived inbred mouse strains from the local NC area, expanding eventually to include other locales.

Conduct toxicity phenotyping on the ~100 strains that this resequencing project has genotyped, paving the way for genotype-phenotype association mapping and genetic analysis.

Develop efficient toxicity indicator phenotypes for a limited set of important environmental chemicals.

Implement whole genome sequencing of Individual mice using the most advanced technology.

Examine laboratory rat strains for variation in genomic DNA sequence.